Design and development of a multimodal biomedical information retrieval system

Dina Demner-Fushman, Sameer Antani, George R. Thoma

National Library of Medicine, Bethesda, MD, USA
{ddemner,santani,gthoma@mail.nih.gov}

Abstract

The search for relevant and actionable information is key to achieving clinical and research goals in biomedicine. Biomedical information exists in different forms: as text and illustrations in journal articles and other documents, in images stored in databases, and as patients' cases in electronic health records. This paper presents ways to move beyond conventional text-based searching of these resources by combining text and visual features in search queries and document representation. A combination of techniques and tools from the fields of Natural Language Processing (NLP), Information Retrieval (IR), and Content-Based Image Retrieval (CBIR) allows developing building blocks for advanced information services. Such services enable searching by textual as well as visual queries, and retrieving documents enriched by relevant images, charts, and other illustrations from the journal literature, patient records and image databases.

1 Introduction

The importance of illustrations in scientific publications is well-established. In a survey of information needs of researchers and educators, Sandusky and Tenopir (2008) found that the scientific journal-article components such as tables and figures are often among the first parts of an article scanned or read by the researchers. In addition, the survey participants indicated that having access to the illustrations [1] prior to obtaining the whole publication will greatly enhance their search experience. In the biomedical domain, Divoli et al. (2010) found that the bioscience literature search systems such as PubMed should show figures from articles alongside search results and that captions should be searched along with the article title, metadata, and abstract. Simpson et al. (2010) showed that retrieval of case descriptions similar to a patient's case was significantly improved with the use of image-related text.

The clear need for a multimodal retrieval system on the one hand, and the sufficient maturity of the IR and CBIR techniques on the other, motivated us to implement a prototype multimodal system (called OpenI) for advanced information services. These services should enable:

- Searching by textual, visual and hybrid queries
- Retrieving illustrations (medical images, charts, graphs, diagrams, and other figures)
- Retrieving bibliographic citations enriched by relevant images
- Retrieving from collections of journal literature, patient records, and independent image databases
- Linking patient records to literature and image databases to support visual diagnosis and clinical decision making

This paper first presents an overview of the processes involved in preparing multimodal scientific articles for indexing and retrieval. It then briefly introduces a distributed architecture that allows for both processing the original documents in reasonable time, and for the real-time retrieval of the processed documents. The paper concludes with the discussion of the implemented system prototype (shown in Figure 1) that currently provides access to over 600,000 figures from over 250,000 medical articles and the future directions in multimodal retrieval.

[1] We use "images", "illustrations", and "figures" interchangeably referring to visual material in our set of medical articles.

Figure 1 Textual search results in the list view of the OpenI system. The navigation panel on the left allows filtering and sorting search results along several facets. The pop-up window that appears on scrolling over the image provides a brief, but key information in the article.

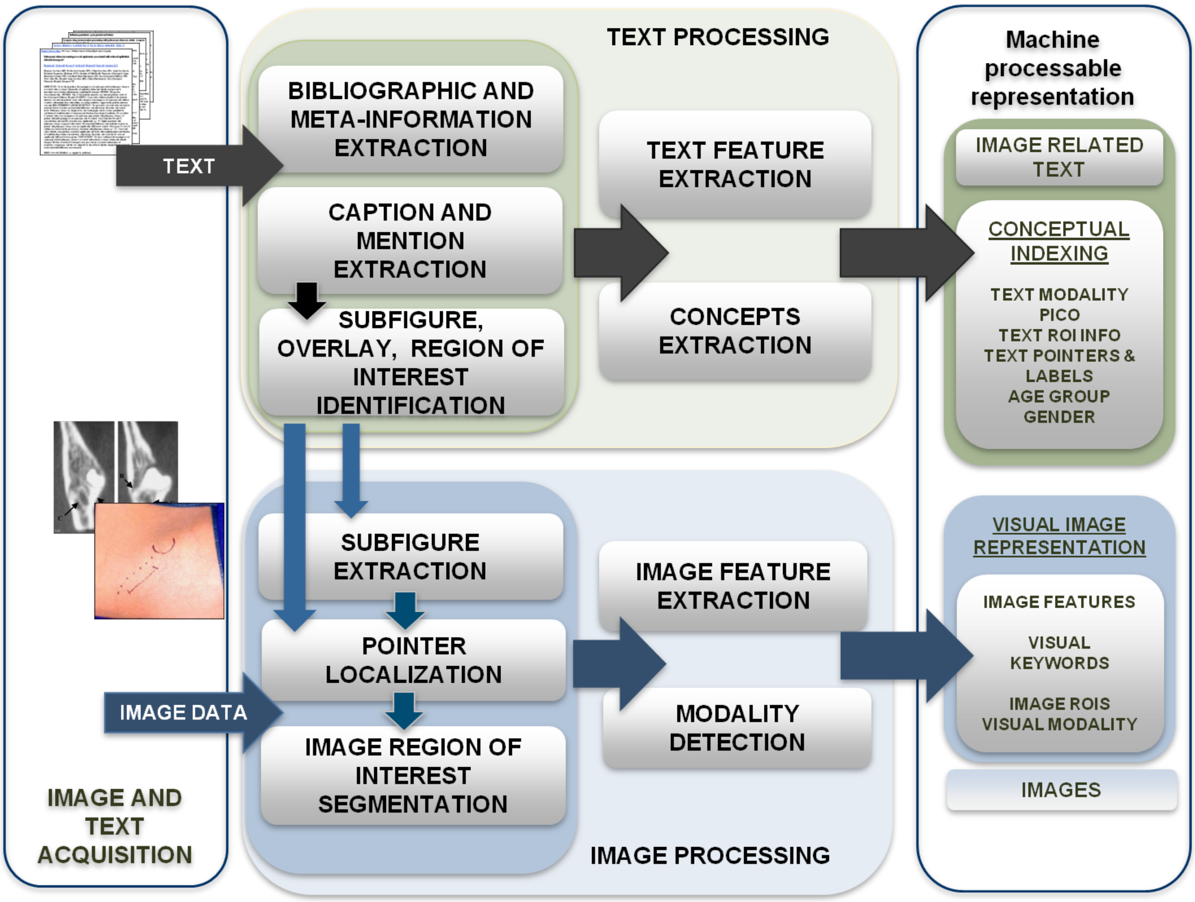

Figure 2 Overview of image and text processing steps for creating enriched citations.

2 Building blocks for advanced information services

To prepare documents for indexing and retrieval, we combine our tools and those publicly available in a pipeline that starts with acquiring data and ends in generation of MEDLINE® citations enriched with image-related information (henceforth, “enriched citations”). The initial separate text and image processing pathways merge in multimodal indexes for use with specialized multimodal information retrieval algorithms as shown in Figure 2. The images and text data in the current prototype are obtained from the open access subset of PubMed Central® (PMC) using the PMC FTP services[2]. This full-text archive of biomedical publications provides the text of each article in XML format and all published figures as JPEG files. The XML files serve as input to the modules assembled in the text processing pipeline, and the images are processed through the image processing pipeline. Several image processing modules require output of the text processing modules as additional input.

The output of the document processing pipeline is a set of enriched MEDLINE citations in XML format that is subsequently indexed with the NLM’s domain-specific search engine Essie (Ide et al., 2007), as well as with the widely-used open-source search engine Lucene[3]. An enriched citation consists of three parts: 1) the original bibliographic citation obtained using E-Utilities[4]; 2) the image caption and image-related paragraphs extracted from the full-text of the article, along with salient information extracted from this free text and stored in structured form; and 3) image features expressed as searchable character strings along with the image URI for display in the user interface.

Some of the processing steps (such as extracting elements of an XML document) are well-known and will be omitted here. We will focus instead on the overall process flow and unique challenges and opportunities presented by images found in biomedical publications.

[2] http://www.ncbi.nlm.nih.gov/pmc/tools/openftlist/
[3] http://lucene.apache.org/
[4] http://www.ncbi.nlm.nih.gov/corehtml/query/static/eutils_help.html

2.1 Generating an enriched citation

The OpenI document processing system is developed in Java and uses Hadoop™ MapReduce to parallelize text processing and image feature extraction. An enriched citation object is generated in the text processing pipeline presented in Section 3. Images are processed independently and the information extracted from the images is added to the enriched citation in the final merging step.

One challenge in image processing arises from several illustrations combined into one figure. These multi-panel (or compound) images found in many articles reduce the quality of image features if the features are extracted from the whole image. For feature extraction, therefore, these images need to be first separated into distinct panels. This process is described in Section 5.

In addition to the text and image features necessary for retrieval, each enriched citation also contains meta-information derived from the basic features (such as the medical terms found in the captions and mapped to the Unified Medical Language System® (UMLS) (Lindberg et al,, 1993) concepts). This meta-information is used to filter and re-rank search results. For example, the results could be restricted to radiology images only, or re-ranked to promote articles focused on genetics. The currently available filters are described in Section 6.

3 Text processing

The text processing begins with extraction of the image caption and the paragraph(s) discussing the figure (“mentions”). In the PMC documents, captions are a defined XML element. We extract the mentions using regular expressions: we first find an indicator that a figure is mentioned, usually, words “Figure” or “Fig” (sometimes within mark-up tags or punctuation) followed by a number, and then extract the paragraph around the indicator.

Next, the caption processing module determines if the caption belongs to a multi-panel figure. The rule-based system is looking for sequences of alphanumeric characters that are included within repeating tags, or followed by a repeating punctuation sign (for example, **A. B. C.**) If a sequence is found, the number of panels and the panel labels are added to the enriched citation.

The next module extracts the descriptions of image overlays (such as arrows) and regions of interest (ROI) indicated by the overlays (Apostolova and Demner-Fushman, 2009). The ROI descriptions added to the enriched citations are currently a searchable field. The whole output of the module is needed for our ongoing research in building a visual ontology that will associate the UMLS concepts with specific image features.


```
  <?xml version="1.0" encoding="utf-8" ?>
- <document>
    <meta iclef_id="239029" />
    <meta publisher="Radiology" />
    <meta journal_title="The puff of smoke sign" />
    <meta fulltext_html_url="http://radiology.rsnajnls.org/cgi/content/full/247/3/910" />
    <meta iti_id="18487544F1" />
    <meta volume="247" />
    <meta authors="Ortiz-Neira, Clara L;" />
    <meta pmid="18487544" />
    <meta issue="3" />
    <title>The puff of smoke sign</title>
    <abstract />
  - <image type="figure" id="1" src="./images/239029.jpg"
      link="http://radiology.rsnajnls.org/cgi/content/full/247/3/910/F1">
      <caption>Anteroposterior angiogram of right internal carotid artery shows abnormal hypertrophy of
        perforating arteries, which produces the puff of smoke sign (arrow) and is associated with
        narrowing (arrowheads) of the M1 and A1 segments of the distal internal carotid artery.</caption>
      <mention />
    - <pico>
        <modalityclass>xr</modalityclass>
        <modality>angiogram</modality>
        <intervention cui="C0002978" negstatus="NOT_NEGATED">angiogram</intervention>
        <anatomy cui="C0226156" negstatus="NOT_NEGATED">right internal carotid artery</anatomy>
        <problem cui="C0020564" negstatus="NOT_NEGATED">hypertrophies</problem>
        <anatomy cui="C1182750" negstatus="NOT_NEGATED">perforating arteries</anatomy>
        <anatomy cui="C0007276" negstatus="NOT_NEGATED">internal carotid artery</anatomy>
      </pico>
    - <rois>
        <roi type="arrow">smoke sign</roi>
        <roi type="arrow">narrowing</roi>
      </rois>
    </image>
  + <mesh>
  </document>
```


Figure 3 Enriched MEDLINE citation

Finally, a concept extraction module submits the captions and mentions to MetaMap (Aronson and Lang, 2010) that identifies UMLS concepts in the text. The module then applies rules and stop-word lists to the MetaMap output to reduce the set of the identified UMLS concepts to the salient disorder, intervention and anatomy terms (Demner-Fushman et al, 2010). Figure 3 shows an enriched citation in XML format.

4 Image processing

Low-level visual features, such as color, texture, and shape, are insufficient for capturing image semantics, but they are the primary building blocks of the visual content in an image. They can be effective if a judiciously selected feature metric is used to capture the visual content in an image, and then incorporated into a suitable machine-learning framework that supports multi-scalar and concept-sensitive visual similarity. In the OpenI prototype system, the low-level visual features extracted from the whole set of images are first clustered and then the resulting clusters are mapped to artificial "words." The words are added to the Enriched citation shown in Figure 3 as the last document preparation step.

4.1 Feature extraction

Images in the open access PMC collection are of different sizes. In order to obtain a uniform measure with greater computational efficiency we compute features from images that are reduced to a common size measuring 256 x 256 pixels. In the future, we intend to process images at a significantly higher (or full) resolution to extract meaningful local features.

Color Features: Color plays an important role in the human visual system and measuring its distribution can provide valuable discriminating data on the image. We use several color descriptors to represent the color in the image. To represent the spatial structure of images, we utilize the Color Layout Descriptor (Chatzichristos and Boutalis, 2008) (CLD) specified by MPEG-7 (Chung et al., 2001). The CLD represents the spatial layout of the images in a compact form and can be computed by applying the discrete cosine transformation (DCT) on the 2D array of local representative colors in the YCbCr color space, where Y is the luminance component and Cb and Cr are the blue and red chrominance components, respectively. Each color channel is 8-bits and represented by an average value computed over 8 x 8 image blocks. We extract a CLD with 10 Y, 3 Cb, and 3 Cr components to form a 16-dimensional feature vector.

Another feature used is the Color Coherence Vector (Pass et al., 1996) (CCV) that captures the degree to which pixels of that color are members of large similarly colored regions. A CCV stores the number of coherent versus incoherent pixels with each color thereby providing finer distinctions than color histograms. Color moments, also computed in the perceptually linear L*a*b* color space, are measured using the three central color moment features: mean, standard deviation, and skewness.

Finally, 4 dominant colors in the standard RGB (Red, Green, Blue) color space and their degrees are computed using the k-means clustering algorithm.

Edge Features: Edges are not only useful in determining object outlines, but their overall layout can be useful in discriminating between images. The Edge Histogram Descriptor (Chatzichristos and Boutalis, 2008) (EHD), also specified by MPEG-7, represents a spatial distribution of edges in an image. It computes local edge distributions in an image by dividing the image into 4 x 4 sub-images and generating a coarse-orientation histogram from the edges present in each of these sub-images. Edges in the image are categorized into five types: vertical, horizontal, 45° diagonal, 135° diagonal, and other non-directional edges. A finer-grained histogram of edge directions (72 bins of 5° each) is also constructed from the output of a Canny edge detection algorithm (Canny, 1986) operating on the image. This feature is made invariant to image scale by normalizing it with respect to the number of edge points in the image.

Texture Features: Texture measures the degree of “smoothness” (or “roughness”) in an image. We extract texture features from the four directional gray-level co-occurrence matrices (GLCM) that are computed over an image. Normalized GLCMs are used to compute higher order features, such as energy, entropy, contrast, homogeneity and maximum probability. We also compute Gabor filters to capture image gist (coarse texture and spatial layout). The gist computation is resistant to image degradation and has been shown to be very effective for natural scene images (Oliva and Torralba, 2006). Finally, we use the Discrete Wavelet Transform (DWT) that has been shown to be useful in multi-resolution image analysis. It captures image spatial frequency components at varying scales. We compute the mean and standard deviation of the magnitude of the vertical, horizontal, and diagonal frequencies at three scales.

Average Gray Level Feature: This feature is extracted from the low-resolution scaled images, where each image is converted to an 8-bit gray-level image and scaled down to 64 x 64 pixels regardless of the original aspect ratio. Next, this reduced image is partitioned further with a 16 x 16 grid to form small blocks of (4x4) pixels. The average gray value of each block is measured and concatenated to form a 256-dimensional feature vector.

Other Features: We extract two additional features using the Lucene image retrieval engine (LIRE) library: the Color Edge Direction Descriptor (CEDD) and the Fuzzy Color Texture Histogram (FCTH) (Lux and Chatzichristos, 2008). CEDD incorporates color and texture information into a single histogram and requires low computational power compared to MPEG-7 descriptors. To extract texture information, CEDD uses a fuzzy version of the five directional edge filters used in MPEG-7 EHD that were described previously. This descriptor is robust with respect to image deformation, noise, and smoothing. The FCTH uses fuzzy high frequency bands of the Haar Wavelet Transform to extract image texture.

5 Multi-panel figure segmentation

Independent of the caption processing module that outputs the number of panels and panel labels, two image-feature based modules detect panel boundaries and panel labels. The output of the three modules is used in the panel splitting module.

The image-feature based panel segmentation module determines if an image contains homogeneous regions that cross the entire image. If no homogeneous regions are found, the image is classified as single-panel. If the homogeneous regions are found, the panel segmentation module iteratively determines if each panel contains homogenous regions, and finally outputs coordinates of each panel.

The label detection module (You et al., 2011) first binarizes the image into black and white pixels, then searches the image for connected components that could represent panel labels, and then applies optical character recognition (OCR) methods to the connected components. Finally, the most probable label sequences and locations are selected from all candidate labels using Markov Random Field modeling.

The label splitting module takes the outputs of the caption splitting, panel segmentation and label detection modules and splits the original figure if the following conditions are met: all three modules agree on the number of panels and the caption splitting and label detecting modules agree on the labels (this happens for approximately 30% of the multi-panel figures). If the panel segmentation or the label detection modules fail completely, the image cannot be split. However, if the modules partially agree on labels and position some of the labels at the corners of the corresponding panels, heuristics help to compensate for the partial errors of individual modules and the combined information helps correctly split another 40% of the multi-panel images.

6 OpenI system architecture

OpenI uses VMware vSphere 4 and a high-performance Linux based SAN to support a fault-tolerant, scalable, and efficient production-grade system. A high-performance SAN storage cluster that was built using Red Hat Enterprise Linux provides OpenI with a dedicated, reliable, predictable and high-performing storage system for the Hadoop cluster. VMware vSphere 4 is used to virtualize the Hadoop Namenode and run it in fault-tolerant mode. The fault-tolerance feature of VMware allows a single virtual machine to run simultaneously on two hardware servers. Further, vSphere monitors the heartbeat of the Namenode and restarts it automatically if it should become inoperable. These features eliminate the single point of failure and turn Hadoop into a stable and reliable development and research platform. Finally, vSphere is used to make OpenI processes run efficiently on multi-core CPUs. Each compute node running processes that are not multi-core aware can be created as a virtual machine tied to a specific physical core.

7 OpenI image retrieval system

The OpenI prototype[5] currently supports image retrieval for textual, visual and hybrid queries. The images submitted as queries are processed as described above and represented using cluster "words". After this processing step, the cluster "words" are treated as any other search terms.

Based on the principles developed by Hearst et al. (2007), the search results are displayed either on a grid that allows a view of all top 20 retrieved images (as shown in Figure 4) or as a traditional list shown in Figure 1. In either layout, scrolling over the image brings out a pop-up window that, along with the traditional elements of search results display, such as titles and author names, provides captions of the retrieved images and short summaries of the articles.

Figure 4 Search results in a grid display

The summaries, which are the "bottom-line" patient-oriented outcomes extracted from

---

[5] http://openi.nlm.nih.gov/

abstracts (Demner-Fushman et al., 2006), are obtained through RIDeM services[6].

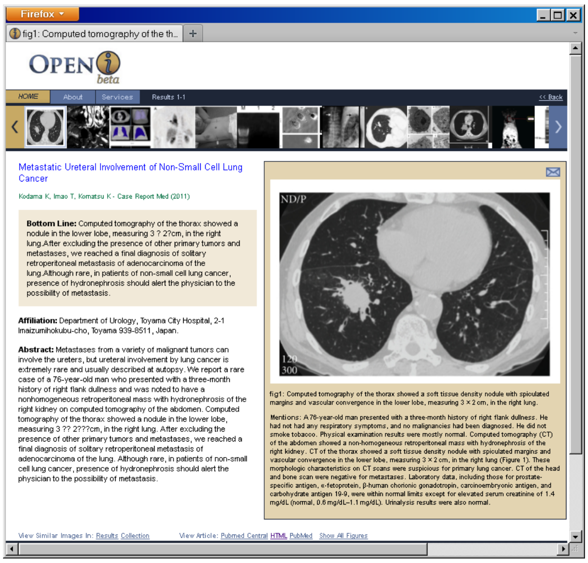

Figure 5 A view of an enriched citation in the user interface. The ribbon at the top allows rapid navigation to other images in the search results. The links at the bottom allow to link out to the publisher's site, PubMed or PubMed Central, and search for similar images. Enriched citations can be sent using the email icon.

Once the search results are displayed, the users can find similar images in the results and in the entire collection. These new searches are based purely on image features. Users may view all other images in a given paper without leaving the initial search page, and drill down to the full enriched citation shown in Figure 5.

The search results can also be filtered using the following facets: 1) image type; 2) subsets; 3) clinical specialties; 4) enriched citation fields.

Image type: The image type filter is based on our classification of images into eight medical images modalities, such as MRI, x-ray, CT, ultrasound and others. Our method (Rahman et al., 2009) uses an SVM to classify images into multiple modality categories. The degree of membership in each category can then be used to compute the image modality. In its basic formulation, the SVM is a binary classification method that constructs a decision surface and maximizes the inter-class boundary between the samples. To extend it to multi-class classification, we combine all pair-wise comparisons of binary SVM classifiers, known as one-against-one or pair-wise coupling (PWC). The PWC method constructs binary SVMs for all possible pairs of classes. Hence, for M classes this method uses M * (M-1) / 2 binary classifiers, each of which provides a partial decision for classifying an image. Each SVM is trained for one image feature. The class with the greatest estimated probability for each feature accumulates one vote. The class with the greatest number of votes after classifying for all features is deemed to be the winning class, and the modality category of the class is assigned to the image. When a user requests a specific image type, a hard constraint on exact match on the image modality field of the enriched citation is imposed.

Subsets: Due to the nature of the collection, not all Medline/PubMed subsets[7], such as the core clinical journals subset[8] are available in OpenI. We used the subject field of the NLM's *List of Journals Indexed in MEDLINE*[9] to categorize the journals into clinical specialties and subsets. Where available, we used the subset field of the original MEDLINE citation.

Enriched citation fields: The users can search the text in any combinations of the following: titles. abstracts, captions, mentions, MeSH terms, and author names.

Finally, the search results could also be re-ranked according to the users' interests along the following axes: 1) by the date of publication (most recent or oldest first; 2) by the clinical task that is discussed in the paper (diagnosis, cause of the problem, prevention, prognosis, treatment, etc.).

8 Related work

Several ongoing research efforts are dedicated to augmenting text results with images. Some of these efforts aim to retrieve images by matching query text terms in the citations to the articles and the figure captions. We list five efforts related to our goals. Most systems do not use image features to find similar images or combine visual and text features for biomedical information retrieval. Our goals include improving relevance of multi-modal (text and

6 http://clinicalreferences.nlm.nih.gov/ridem/

7 http://www.nlm.nih.gov/bsd/pubmed_subsets.html

8 http://www.nlm.nih.gov/bsd/aim.html

9 http://www.nlm.nih.gov/services/journal_list.html

image) information retrieval by including lessons learned from these efforts.

The BioText (Hearst et al., 2007) search engine, searches over 300 open access journals and retrieves figures as well as text. BioText uses Lucene to search full-text or abstracts of journal articles, as well as image and table captions. Retrieved results (displayed in a list or grid view) can be sorted by date or relevance. This search engine has influenced our user interface design.

Yottalook [10] allows multilingual searching to retrieve information (text or medical images) from the Web and journal articles. The goal of the search engine is to provide information to clinicians at the point of care. The results can be viewed as thumbnails or details. This site sets an example in the breadth of its searches, capabilities to filter results on image modality and other criteria, being current with social media, and connecting with the users' myRSNA accounts (offered by the Radiological Society of North America -- RSNA) that allows saving search results.

Other related work includes the Goldminer [11] search engine developed by the American Roentgen Ray Society (ARRS) that retrieves images by searching figure captions in the peer-reviewed journal articles appearing in the RSNA journals Radiographics and Radiology. It maps keywords in figure captions to concepts from the UMLS. Users have the options to search by age/modality/sex for images where such information is available. Results are displayed in a list or grid view.

The FigureSearch [12] system uses a supervised machine-learning algorithm for classifying clinical questions and Lucene for information retrieval over the published medical literature to generate a list view of the results with relevant images, abstracts, and summaries.

The Yale Image Finder (YIF) (Xu et al., 2008) searches text within biomedical images, captions, abstract, and title to retrieve images from biomedical journal papers. YIF uses optical character recognition to recognize text in images in both landscape and portrait modes.

The IRMA system [13] primarily uses visual features and a limited number of text labels that describe the anatomy, biosystem, the imaging direction, and modality of the image for medical image retrieval. We have collaborated with the developers of the IRMA system, and enhanced their image retrieval system (that uses features computed on the gross image) with our image features and similarity computation techniques applied to local image regions (Antani et al., 2007).

Increasing commercial interest in multi-modal information retrieval in the biomedical domain is indicated by the industry teams participating in the ImageCLEFmed [14] evaluations dedicated to retrieval of medical images and similar patients' cases. Participants include researchers from Siemens, GE Medical Systems, Xerox, and other industrial organizations. Publishers such as Springer also provide text-based image retrieval [15]. Other commercial image search engines include those developed by Google, Gazopa, and Flickr.

9 Future work

The deployment of the prototype system and the architecture presented in this paper allows continuing research in several directions. First, we are interested in the usability of the current user interface and the usefulness of the search features. Second, we are expanding and improving the extraction of the basic image features and the selection of these features for various higher-level tasks, such as image modality classification, image ROI recognition, and building a visual ontology. The latter tasks includes associating specific image features with the UMLS concepts. Recognizing that many researchers would like to focus on specific aspects of image retrieval, for example, improving retrieval methods, or text understanding, we plan to provide our current document processing methods as publicly available services.

[10] http://www.yottalook.com/
[11] http://goldminer.arrs.org/home.php
[12] http://figuresearch.askhermes.org/articlesearch/index.php?mode=figure

[13] http://www.irma-project.org
[14] http://www.imageclef.org/2012/medical
[15] http://www.springerimages.com/

Acknowledgments

This work would have not been possible without our amazing team of post-doctoral researchers and the equally talented OpenI staff. Our thanks go to Emilia Apostolova, Michael Chung, Glenn Ford, Michael Kushnir, Srinivas Phadnis, Md. Rahman, Matthew Simpson, Daekeun You and Zhiyun Xue. We thank Essie developers Russell Loane and Nicholas Ide for providing the search engine and being wonderfully responsive and supportive in development of the OpenI systems.
This work was supported by the Intramural Research Program of the National Library of Medicine, National Institutes of Health.

References

Antani SK, Deserno TM, Long R, Thoma GR. Geographically Distributed Complementary Content-Based Image Retrieval Systems For Biomedical Image Informatics. Proc. MEDINFO 2007. 12(1):493-7.

Apostolova E, Demner-Fushman D. Towards Automatic Image Region Annotation - Image Region Textual Coreference Resolution. In Proceedings of the 2009 Annual Conference of the North American Chapter of the Association for Computational Linguistics. June 2009. Boulder, Colorado. Association for Computational Linguistics, p 41--44.

Aronson AR, Lang FM. An overview of MetaMap: historical perspective and recent advances. J Am Med Inform Assoc. 2010 May-Jun;17(3):229-36.

Canny J. A Computational Approach To Edge Detection, IEEE Trans. Pattern Analysis and Machine Intelligence, 8(6):679–698, 1986.

Chatzichristos SA, Boutalis YS. CEDD: Color and edge directivity descriptor: A compact descriptor for image indexing and retrieval. Proc. 6th International Conference on Computer Vision Systems, Lecture Notes in Computer Science, 5008:312-22, 2008.

Chang SF, Sikora T, A. Puri A. Overview of the MPEG-7 standard. IEEE Transactions on Circuits and Systems for Video Technology, 11(6):688-95, 2001.

Demner-Fushman D, Few B, Hauser SE, Thoma G. Automatically identifying health outcome information in MEDLINE records. J Am Med Inform Assoc. 2006 Jan-Feb;13(1):52-60.

Demner-Fushman D, Mork JG, Shooshan SE, Aronson AR. UMLS content views appropriate for NLP processing of the biomedical literature vs. clinical text. J Biomed Inform. 2010 43(4): 587-594

Divoli A, Wooldridge MA, Hearst MA (2010) Full Text and Figure Display Improves Bioscience Literature Search. PLoS ONE 5(4): e9619.

Hearst MA, Divoli A, Buturu H, Ksikes A, Nakov P, Wooldridge MA, et al. Biotext search engine: beyond abstract search. Bioinformatics 2007;23(16):2196–7.

Ide NC, Loane RF, Demner-Fushman D. Essie: a concept-based search engine for structured biomedical text. J Am Med Inform Assoc. 2007 May-Jun;14(3):253-63.

Lindberg DAB, Humphreys BL, McCray AT. The Unified Medical Language System. Meth. Inform. Med. 1993; 32: 281-291

Lux M, Chatzichristos SA. LIRe: Lucene image retrival: an extensible java CBIR library. Proc. 16th ACM International Conference on Multimedia, 2008,1085-88.

Oliva A, Torralba A. Building the Gist of a Scene: The Role of Global Image Features in Recognition. Progress in Brain Research. Visual Perception - Fundamentals of Awareness: Multi-Sensory Integration and High-Order Perception. 155, Part 2, 2006:23-36

Pass G, Zabih R, Miller J. Comparing images using color coherence vectors. Proc. 4th ACM International Conference on Multimedia, pages 65-73, 1996.

Rahman MM, Antani SK, Long LR, Demner-Fushman D, Thoma GR. Multi-Modal Query Expansion Based On Local Analysis For Medical Image Retrieval. Lecture Notes in Computer Science. First MICCAI International Workshop on Medical Content-Based Retrieval for Clinical Decision Support (MCBR-CDS 2009); (doi: 10.1007/978-3-642-11769-5):110-9.

Sandusky RJ, Tenopir C. Finding and Using Journal Article Components: Impacts of Disaggregation on Teaching and Research Practice. Journal of the American Society for Information Science & Technology, 59 (6) April 2008: 970-82.

Simpson MS, Demner-Fushman D, Thoma GR. Evaluating the Importance of Image-related Text for Ad-hoc and Case-based Biomedical Article Retrieval. AMIA Annu Symp Proc. 2010 Nov 13;2010:752-6.

Xu, S., McCusker, J., Krauthammer. M. (2008). Yale Image Finder (YIF): a new search engine for retrieving biomedical images. Bioinformatics, 2008

You D, Antani S, Demner-Fushman D, Govindaraju V, Thoma GR Detecting figure-panel labels in medical journal articles using MRF. Proceedings of 11th International Conference on Document Analysis and Recognition (ICDAR 2011) (pp. 967-971). Beijing, China.